1. The authors provide sufficient details for most of the analysis. However, some methods are presented with insufficient details.

2. All the results that support the conclusion are directly shown.

3. I find that the conclusion drawn by the authors is an exaggerated extension of the results obtained. I think a major flaw is present in the definition of ‘states’, which are the core of the study. Clear definition of the term ‘state’ is needed, together with additional analysis to show their existence. At this stage, the results are too weak to support the conclusion.

4. The study complies with the ethical guidelines.

**Major comments:**

* Defining ‘encoding’ and ‘retrieval’ states requires a careful, complete, and clear analysis. From Fig.2, it seems that NT points are scattered and do not have stable ‘states’. This could be a major flaw of the study, since the ‘transition between states’ could be no longer valid. To define states, authors would need to explicitly show that points that belong to ‘encoding’ and ‘retrieval’ phases are clustered in different regions of the 3-dimensional space and have small overlap.
* The results presented in Fig.7 need to be compared with the cosine distance distribution of random vectors in the 3-dimensional space. Among the ones shown in Fig.7, which are the significant results?
* From Fig.1, the encoding NT seems to be close to 0 the whole time, apart from when SWR+ appears. To validate this switching between encoding and retrieval states, the authors would need to compare the cosine similarity distribution of rSWR+ with gEgR, and the distribution of rSWR+ with gR. If the authors find no significant difference in the two distributions, then they cannot claim the switching between encoding and retrieval states.
* The authors do not provide sufficient details on the SWR clustering with UMAP. Please write how you define the silhouette score and which features do you use for clustering.
* I find that section 3.6 and Fig.6 are not adding any information to the study. The results obtained with the initial 3-dimensional projection provide the same results as the ones in section 3.6.
* The authors find positive correlation between set size and the distance between median NT position in encoding (gE) and median NT position in retrieval (gR). Distances are log-transformed. Is this correlation present also when distances are not log-transformed? I find that the reason presented in Discussion for log-transformation (‘log-normal distributions are prevalent in the central nervous system’) is not valid. I think that this Discussion paragraph should be deleted.

**Minor comments:**

* Is the result shown in Fig.2c from a single patient?
* Fig.3c is not informative. By definition, SWR+ and SWR- will have the same duration distribution.
* Contrary to what is written in the Figure caption, Fig.4d shows the mean SWR incidence, but not the confidence interval. In addition, the SWR- incidence is not present.
* In Fig.3 the authors do not use SWR information. The reference to Fig.3 in section 3.6 is not correct.
* In Discussion, the phrase ‘hippocampal neurons form unique NTs, primarily during SWR’ is not correct. Hippocampal activity always forms NT. Please remove or rephrase.
* In Discussion, the phrase ‘the potential bias does not substantially challenge our main findings’ is not clear. Why does the potential bias do not affect the main findings?